NLSY97 Appendix 6:

Event History Creation and Documentation

The NLSY97 survey records significant life-course transitions experienced by young people, such as education, employment, program participation, and marital history, in a longitudinal format. The event history arrays document these events in a chronological format that records the significant transitions in a meaningful manner while maintaining data quality. Using these arrays, researchers can extract the status of a respondent at a point in time or over time. Event history arrays are generated for four distinct areas: employment, marital/cohabitation status, program participation, and schooling. This section presents information on each type of event history array; for details on the chronological format of the arrays and the naming conventions used to identify the variables, users should refer to appendix 7.

Employment Event History Arrays

Three employment arrays provide information on the respondent's employment on a weekly basis. These arrays include information about employer jobs only; freelance jobs were not included in the arrays. All employment arrays provide information starting in the week that the respondent turned 14 and ending in the week that he or she was last interviewed.

1. EMP_STATUS

This main array presents the employment status of a respondent in a particular week. The codes and their explanations follow:

Code	Definition
Status=0: No information reported to account for week	Not assigned during round 1.
Status=1: Not associated with an employer, not actively searching for an employer job	Refers to weeks during a between-jobs gap in which the respondent is not actively searching and reports working at a freelance job. Since the actual weeks working at a freelance job cannot be determined, all weeks in which the respondent is not actively searching are coded in this manner. Similar information was not collected for the within-job gaps during round 1.
Status=2: Not working (unemployment vs. out of labor force cannot be determined)	Assigned when the respondent is not asked follow-up questions about his or her search activity during a within-job gap or a between-jobs gap.
Status=3: Associated with an employer, periods not working for employer are missing	Used when a respondent reports an indeterminate start or stop date for a within-job gap.
Status=4: Unemployed	Indicates that the respondent reports actively searching for work during a within-job gap or a between-jobs gap. When the number of weeks unemployed do not account for the entire gap period, weeks unemployed are assumed to occur in the middle of that period.
Status=5: Out of the labor force	Assigned during a between-jobs gap or a within-job gap when the respondent is either not actively searching for work or on layoff from a job.
Status=6: Active military service	Indicates that the respondent is tied to the military.
Status=9701 to 9809: Employer on roster	Refers to the employer number on the employer roster (YEMP_UID.01 to YEMP_UID.09). Presence of an employer number indicates that the respondent was working during a given week. Civilian work takes precedence over other activities, such as job search. Respondents who report working at an employer job for one day in a given week are listed as having worked at that job for the entire week, regardless of other activities.

2. EMP_DUAL_JOB#
If a respondent holds more than one employer job during a week, the second employer job is presented in a dual job array. These arrays contain only the job number of the overlapping job; labor force status information is only included in the main array. For example, if a respondent held two employer jobs (e.g., the first and third jobs listed on the employer roster), during the 52nd week of 1997, the employer number for the first job would be recorded in the EMP_STATUS array and the employer number for the third job would be recorded in the EMP_DUAL_2 array. If a respondent held three jobs (e.g, jobs #01, #04, and #05 on the roster) in one week, the first job would be recorded in the EMP_STATUS array, the employer ID for job #04 would be recorded in the EMP_DUAL_2 array, and the employer ID for job #05 would be recorded in the EMP_DUAL_3 array.

3. EMP_HOURS
This final array calculates the total number of hours worked by a respondent at any employer job during a given week. Hours per week worked at each job are assumed constant except during a reported gap, when the hours for that job are assumed to be zero. Each week is assigned a code of '-3 (invalid skip)' when any of the jobs has an indeterminate gap date.

A secondary set of variables translates the reported beginning and ending dates (day, month, and year) of employer jobs and the gaps within those jobs to the week and year naming scheme (e.g., EMP_GAP_START_YEAR.01.01 and EMP_GAP_END_YEAR.01.01 provide the start and end dates of the respondent's first gap in the continuous week and year format). More information about the week and year naming scheme is provided in appendix 7 in this document.

The created event history variables can be used in conjunction with the main file information about the respondent's employment. The 9701–9809 codes that appear in the status and dual job arrays are unique employer ID numbers that are also used in the main file data. In the main data, these codes are listed in the YEMP_UID.xx variables (e.g., R24761.). Using these unique ID codes, researchers can identify the comparable job information (e.g., start and stop dates, fringe benefits, job satisfaction, industry and occupation, etc.) from the main file. Employers first reported in the round 1 interview have codes beginning with "97"; new employers in round 2 were assigned codes beginning with "98." This system permits users to link employers across survey rounds and across data files and to identify the round in which an employer was first reported.

Marital Status Event History Arrays

The NLSY97 marital and cohabitation arrays record changes in the respondent's marital status and cohabitation changes on a monthly basis. The marital/cohabitation history program converts dates reported in the marriage section (beginning and ending dates of cohabitations, marriages, separations, divorces, and widowhoods) to an actual month number, using January 1980 as month #1. Used jointly, these arrays allow the researcher to obtain a detailed history of the respondent's partners and changes in his/her marital and cohabitation status on a monthly basis. All marital/cohabitation arrays provide information beginning in the month that the respondent turned 14 and ending in the month that he or she was last interviewed. Additionally, the beginning dates of the youth's first marriage and first cohabitation are provided in two variables: CV_FIRST_MARRY_MONTH and CV_FIRST_COHAB_MONTH.

Four types of arrays record transitions between living without a partner of the opposite sex to cohabiting or to marriage.

1. MAR_STATUS
The main array presents the status (e.g., never married/not cohabiting, cohabiting, married, divorced) of a respondent during a particular month. Marital status takes precedence over cohabiting; for example, if a

respondent is divorced and living with another partner, the status listed in this array will be 'divorced.' Respondents who are married but not living with their spouse are coded as married. When a respondent reports an annulment, the marriage is maintained and the marital status code after the annulment is 'divorced.'

2. MAR_COHABITATION
This second array details the partner that the respondent is living with in a particular month. For example, if the respondent is cohabiting, the variable for each month identifies whether the respondent lives with partner 1, partner 2, spouse 1, spouse 2, etc. Users should note that "1" and "2" in this case refer to the respondent's partners/spouses in chronological order. The numbers do not necessarily refer to the same person as the spouse/partner questions asked directly of the respondent during the survey.

If a respondent reports living with two partners in the same month, the first partner is listed in this array, subsequent partners are listed in the MAR_DUAL array.

3. MAR_DUAL
If there is an overlap of partners (e.g., partner 1 leaves at the beginning of the month and partner 2 moves in at the end of the month), this array records the presence of the new partner. The format of these variables is the same as that of the MAR_COHABITATION variables.

4. MAR_PARTNER_LINK
The fourth array links the cohabiting partner or spouse to the partner order in the main survey questions. For example, a number of codebook variables report the status of "Partner #01" or "Partner #02." If the person the respondent reports cohabiting with in a given month is the first person asked about in the 1997 survey, the MAR_PARTNER_LINK variable for that month is 9701. If the partner in a given month is the second partner asked about in 1997, this variable is coded 9702, and so on. This array allows the researcher to identify characteristics of the respondent's partner and to link them with spells of marriage or cohabitation. In round 2, if a partner was reported in round 1, he or she retains the same identification number. Otherwise, new partners are coded as 9801, 9802, etc., in the same manner that was used in round 1.

Program Participation Event History Arrays

Program participation arrays are constructed individually for five programs—Worker's Compensation, Unemployment Compensation, AFDC, Food Stamps, and WIC. The AFDC array includes all federal and state programs created under Temporary Assistance to Needy Families (TANF) or any government program for needy families that replaces AFDC. All other programs (e.g., LIHEAP, SSI, other) are combined into a sixth array entitled 'Other.' For each program type, except Worker's and Unemployment Compensation, three arrays are created. All program participation arrays provide information starting in the month that the respondent turned 14 and ending in the month that he or she was last interviewed.

A secondary set of variables translates the reported beginning and ending dates (month and year) of a spell within the program into the continuous month scheme (e.g., AFDC_START_MONTH and AFDC_STOP_MONTH). More information about the continuous month scheme is provided in appendix 7 in this document.

1. STATUS
The main array, (e.g., AFDC_STATUS), presents the status—receiving or not—of a respondent during each month. When asked for the start or stop date of a spell, the respondent could respond 'don't know' or 'refuse' to any component. In this case, the respondent was then asked how many weeks the spell lasted. The number of reported weeks was then divided by 4.3 to determine the equivalent number of months. If a fraction of a month was reported, then the entire month was counted as a month receiving

benefits. Using a combination of start date, stop date, and week information, each spell was defined and a value of '1' inserted into the status array to indicate months of receipt. The months that a respondent did not receive that benefit, but was eligible to receive it, have a value of '0.' An edit variable, (e.g., AFDC_EDIT_DATE) flags respondent-reported and imputed dates. The process by which imputed dates and the corresponding edit flag were assigned is described below:

Flag	Definition
Edit Flag=1: Respondent reported participation dates	Respondent reported a complete start and stop date and is not currently receiving. If the respondent reports still receiving at the time of the interview, the interview date is assigned as the temporary stop date. In the next survey round, the respondent will be asked if he or she is still receiving; if not, a permanent stop date equivalent to the previous round's interview date will be assigned. If the respondent reports receiving, participation will continue in filling the array.
Edit Flag=2: Start month imputed	**Total weeks known:** If the respondent reports not currently receiving, then set the month equal to January and count forward by the number of weeks to imply a stop date. If currently receiving, then count back by the number of weeks from the interview date to impute a start month. If the month indicated by the count falls short of the start year, the start month is December of the start year. If the month occurs in the year before the reported start year, then the start month is January of the start year.
	Total weeks unknown: If the respondent reports not currently receiving, then the start month is set to January. Use December as the stop month and the start year as the stop year. If the respondent reports currently receiving, use December as the start month.
Edit Flag=3: Start month and year imputed	**Total weeks known:** Count back by the number of weeks from the interview date if currently receiving. If not currently receiving, then count back from interview date to find the most recent year the respondent could have begun receiving and call the start date January of that year; then count forward the number of weeks from that date to imply a stop date.
	Total weeks unknown: If currently receiving, begin the spell at the respondent's 14th birthday.
Edit Flag=4: Stop month imputed	**Total weeks known:** If not currently receiving, then count forward from start date. If the month indicated falls short of the stop year, then use January of the stop year as the stop month; if the number of months exceeds the stop year, then set the stop month to December of the stop year. If the stop year is equal to the interview year and the stop month exceeds the interview month, then stop at the interview date.
	Total weeks unknown: If not currently receiving, then use December of stop year for the stop month.
Edit Flag=5: Stop month and year imputed.	**Total weeks known:** If not still receiving, count forward from the start date.
	Total weeks unknown: If not currently receiving, then use December of the start year as the stop month and the start year as the stop year.
Edit Flag=6: Start and stop dates imputed.	**Total weeks unknown:** The imputed dates are based on the previous interview's date (start date) to the current interview date (stop date); in round 1, the last interview date is the respondent's 14th birth month and year.
Edit Flag=7: Start and stop dates complete but gap information missing.	In these cases, the respondent began a spell of receipt in round 1 and ended it in round 2. The start and stop dates are accurate, but no information was collected about gaps in receipt.

2. AMOUNT RECEIVED

If a respondent reports receiving in a particular month, a second array presents the amount received in each month (e.g., AFDC_AMT). The dollar values asked about during the interview were meant to be monthly values. However, some responses were higher than the federal or state limits on the amount

received from a particular benefit. A likely reason is that the respondent mistakenly reported a total value rather than a monthly value. Values determined to be too high were divided by the number of months the respondent reported receiving the benefit. These values were used in the AMT arrays instead. A second set of edit variables (e.g., AFDC_EDIT_AMT) flags these values for a particular spell. In round 1, the edit values were set for the latest year available. For AFDC, the highest benefit offered was $1229; reported amounts falling above this maximum were edited. The maximum amount accepted for food stamps was $936; the edit flag indicates reported amounts that fall above this maximum.

3. HOUSEHOLD MEMBERS RECEIVING
If a respondent reports receiving in a particular month, the persons in the household who benefit from the program in each month (e.g., respondent only, child only, respondent and child) are recorded in a third array (e.g., AFDC_HH). This program condenses the set of answers from the question in the survey that collects this information; for example, see YPRG-18300.01_001 to YPRG-18300.01_005 for AFDC. Users should note that Worker's Compensation and Unemployment Compensation are not included in this array because these programs are collected for the respondent only.

IV. Schooling Event History Arrays

Yearly Schooling Variables: A set of schooling variables provides information for each year beginning in 1980, the year when the first information is available in the survey, through 1998. These education arrays are somewhat different than the other event history arrays. Information on a respondent's education is reported on a yearly basis, rather than monthly or weekly. This approach is used to combine information from the youth questionnaire, which collects more detailed data, and from the round 1 parent questionnaire, which presented information only for each year. In general, these variables refer to the school year rather than the calendar year. That is, 1991 in a variable title or in the data for a variable generally indicates the school year starting in fall 1991 and ending in spring 1992.

Users should be aware that, because questions were not identical in the round 1 parent questionnaire and the round 2 youth questionnaire, the transition between the two data sources was not seamless and some information for the yearly variables had to be imputed. If they feel that a given value is questionable, researchers may wish to compare created variables to the raw data and to the monthly schooling arrays described below.

1. SCH_YEAR_TO_GRADE
This array presents the grade the respondent attended during the school year. The last two digits of the question name indicate the school year. For example, SCH_YEAR_TO_GRADE.90 refers to the grade attended by the respondent during the school year that starts in fall 1990 and ends in spring 1991.

2. SCH_GRADE_TO_YEAR
This array refers to the year the respondent attended a certain grade. For example, if the respondent attended second grade in 1992–93, then SCH_GRADE_TO_YEAR.2 would have the value 1992.

3. SCH_CHANGES
This array counts the number of times the respondent changed the school attended during the school year. For example, SCH_CHANGES.90 shows how many different schools the respondent attended during the school year that started in fall 1990 and ended in spring 1991.

4. SCH_MNTHS_MISSED
This array presents the number of months during the school year that the respondent did not attend school. For example, if SCH_MNTHS_MISSED.90 has a value of 3 for a respondent, then that respondent had a gap in attendance of three months during the school year that started in the fall of 1990

and ended in the spring of 1991. A gap is defined as missing school for one or more months (not including summer vacation); gaps do not have to be consecutive.

5. SCH_SUMMER_SCHOOL
This array refers to extra school classes during an educational break in a given school year, such as summer school. For example, SCH_SUMMER_SCHOOL.90 shows whether the respondent attended school during a break in the 1990–91 school year.

6. SCH_GRADE_PROGRESS
This array has positive values if there are any special events that occurred during the school grade. For example, a positive value in SCH_GRADE_PROGRESS.2 indicates that the respondent was skipped or demoted during second grade. Researchers should note that parents might have been confused as to how to answer the skip grade questions asked during the interview. For example, there are parents who say their child skipped from 5th to 6th grade, while others say from 4th to 6th grades. Both of these cases are probably stating that the child missed most or all of the 5th grade. To resolve this ambiguity, the code states that if a child is skipped consecutive years then the first year (i.e. 5th grade) was missed. If a parent reports non-consecutive years (i.e. 4th to 6th) then the program assumes the year(s) in the middle are the ones not attended.

7. SCH_YEAR_PROGRESS
This array refers to any special events that occurred during the school year. The question name's last two digits indicate the school year this variable refers to. For example, SCH_YEAR_PROGRESS.90 shows special events that occurred during the school year that starts in fall 1990 and ends in spring 1991. The special events, such as grades skipped or demoted to, are defined in the same way as in the previous array.

8. SCH_SUSPENSIONS
This array counts the number of days during the school year the respondent was suspended from school. For example, if SCH_SUSPENSIONS.90 has a value of 3 then the respondent was suspended from school 3 days during the school year that started in fall 1990 and ended in spring 1991.

Monthly Schooling Variables. In round 2, 3 types of monthly arrays and one flag variable (SCH_DUAL_1998) were created. Each array captures information for each month from the respondent's interview date in round 1 to the interview date in round 2.

1. SCH_STATUS
This array reports the respondent's enrollment status during each month from the round 1 interview date through the current interview date. Coding categories include unknown, not enrolled, in grades K to 12, in college, on vacation, expelled, and other.

2. SCH_TERM
These variables report the respondent's school type and grade for each month in the time period. The first two digits represent the type of school (public = 10, private = 20, and religious = 30). The last two digits provide the respondent's grade in school (1–12) or year in college (1–8). To determine whether the school is a K–12 school or a college, researchers are advised to combine this variable with the SCH_STATUS variable described above.

3. SCH_ID
This variable permits users to link array information to the school roster in the main data file and access other information about the school. The first two digits represent the survey year the respondent first reported attending the school. The last two digits provide the number of the school on the current survey year's roster. For example, a value of 9701 indicates that the school was first reported in round 1 and is school #01 on the round 2 school roster.

4. SCH_DUAL_1998

A small number of NLSY97 respondents went to two different schools in the same month. Because only the first school can be reported in the other arrays, this variable flags these special cases. There is only one variable for each school for the period between the round 1 and 2 interviews; the exact month when the overlap occurred is not indicated, and overlap may have occurred in more than one month.

Flag	Definition
Flag=1	More than one K–12 school attended in the same month
Flag=2	More than one college attended in the same month
Flag=3	More than one K–12 school attended in the same month AND More than one college attended in the same month
Flag=4	K–12 school and college attended in the same month
Flag=5	More than one K–12 school attended in the same month AND K–12 school and college attended in the same month
Flag=6	More than one college attended in the same month AND K–12 school and college attended in the same month
Flag=7	More than one K–12 school attended in the same month AND more than one college attended in the same month AND K–12 school and college attended in the same month